3. Key limitations of the available epidemiological studies on pesticides

3.1. Limitations identified by the authors of the EFSA external scientific report

The EFSA External scientific report (Ntzani et al., 2013, summarized in Annex A) identified a plethora of epidemiological studies which investigate diverse health outcomes. In an effort to systematically appraise the epidemiological evidence, a number of methodological limitations were highlighted including the lack of direct exposure assessment, use of generic pesticide definitions, multiple testing, and heterogeneity of data. In the presence of these limitations, robust conclusions on causality based on epidemiological evidence alone could not be drawn, but outcomes for which supportive evidence from epidemiology existed were highlighted for future investigation. The main limitations identified included:

Weak study designs: Lack of prospective studies and frequent use of study designs that are prone to bias (recall bias and reverse causation for case-control and cross-sectional studies). In addition, many of the studies conducted appeared to be insufficiently powered.

Lack of detailed exposure assessment, including lack of appropriate biomarkers. Instead many studies relied on broad definition of exposure assessed through questionnaires (often not validated). There was often also lack of information on specific pesticide exposure and co-exposures.

Deficiencies in outcome assessment (broad outcome definitions and use of self-reported outcomes or surrogate outcomes).

Deficiencies in reporting and analysis (interpretation of effect estimates, confounder control and multiple testing).

Selective reporting, publication bias and other biases (e.g. conflict of interest) were likely to be prevalent in this literature.

In many cases the quality of the studies was suboptimal, and for many health outcomes too few studies were available. The observed heterogeneity in the results within each studied outcome was often large. However, heterogeneity is not always a result of biases and may be genuine and consideration of a priori defined subgroup analysis and meta-regression should be part of evidence synthesis efforts. Occupational studies, which are of particular importance to pesticide exposure, are also vulnerable to the healthy worker effect, a bias resulting in lower morbidity and mortality rates within the workforce than in the general population. The healthy worker effect tends to decline with age of the population under study.

Good-quality studies with sufficient statistical power, detailed definition of pesticide exposure and transparent reporting are rare. Apart from the Agricultural Health Study, there were no other large studies with good quality data for many study outcomes. It is important to note that several of these methodological limitations have not been limited to pesticide exposure studies and, most importantly, are not specific in epidemiology and have been observed in other specific fields including in animal studies (Tsilidis et al., 2013).

Given the wide range of pesticides with various definitions in the EFSA External scientific report, it is difficult to harmonise this information across studies. Although heterogeneity of findings across studies can be as informative as homogeneity, information needs to be harmonised such that replication can be assessed and summary effect sizes be calculated. This does not mean that if there is genuine heterogeneity the different studies cannot be pooled. Limited conclusions can be made from a single study. Nonetheless, the report highlighted a number of associations between pesticides and health effects that merit further consideration and investigation. Of interest is the fact that a considerable proportion of the published literature focused on pesticides no longer approved for use in the EU and in most developed countries e.g., studies focusing solely on DDT and its metabolites constituted almost 10% of the eligible studies (Ntzani et al., 2013). These may still be appropriate since they may persist as pesticide residues or because they continue to be used in developing countries. Also, the report focused on epidemiological evidence in relation to any health outcome across a 5-year window. Although the report is valuable in describing the field of epidemiological assessment of pesticide-

health associations, it is not able to answer specific disease-pesticide questions thoroughly. A more in-depth analysis of specific disease endpoints associated with pesticides exposure is needed where this information is available and studies published earlier than the 5-year window should be also included.

3.2. Limitations in study designs

For ethical reasons randomized controlled trials are not generally allowed to test the safety of low dose pesticide exposure in the EU. Therefore, information on potential adverse health consequences in humans has to be extracted using observational studies. Ideally such studies should be prospective and designed so that the temporal separation between the exposure and the disease outcome is appropriate with respect to the time it takes to develop the disease. For outcomes such as cancer or cardiovascular diseases, which often have a long latency period (>10 years), exposure should be assessed more than once prior to the outcome assessment. Exposure at one time point may not accurately reflect long-term exposure. The problem is that the disease may not have been identified at the time of the exposure assessment so reverse causality is a problem. For this reason, sometimes the outcomes identified during the first 2 years of follow-up need to be excluded. For other outcomes with a shorter latency period such as immune function disturbances the appropriate temporal separation may be in the range of days or weeks and a single exposure assessment may be adequate. In short, the ideal design of a study depends on the latency period for the outcome under consideration. The expected latency period then determines both the length of follow-up and the frequency for which the exposure has to be quantified. Failure to consider these issues when designing a study means that the exposure and outcome cannot be reliably linked.

Among the 795 studies reviewed in the Ntzani report 38% were case-control studies and 32% cross-sectional studies. As a result, evidence on potential adverse health consequences of pesticide exposure is largely based on studies that have sub-optimal design, at least for outcomes that have long latency periods. For the cross-sectional studies, directionality cannot be assessed and observed associations may often reflect reverse causation (is the disease caused by the exposure, or does the disease influence the exposure?). However for pesticides reverse causation could be observed.

Although case-control studies are frequently used for rare outcomes, such as several cancers, their main limitation is that they are prone to recall bias and they have to rely on retrospective assessment of exposure. Alone, case-control studies generally provide rather weak evidence, but they can still provide useful information, especially for rare outcomes. It is important to examine whether results from case control and prospective studies converge. This was for example the case amongst studies that were conducted to examine associations between intake of trans-fatty acids and cardiovascular disease (EFSA 2004), where both case-control and prospective studies consistently reported positive associations. The effect estimates between the two study designs were systematically different with prospective studies reporting more modest effect sizes but both study designs reached similar conclusions.

3.3. Relevance of study populations

Because the environmentally relevant doses of pesticides to which individuals are exposed are lower than those required to induce observed toxicity in animal models, the associated toxic effects need to be understood in the context of vulnerable subpopulations. This is the case of genetic susceptibility, which represents a critical factor for risk assessment that should be accounted for (Gómez-Martín et al., 2015).

One other subgroup of population of special interest are represented by children, because their metabolism, physiology, diet and exposure patterns to environmental chemicals differ from those of adults and can make them more susceptible to their harmful effects. The window(s) of biologic susceptibility remain unknown for the most part, and would be expected to vary by mechanism. Those subgroups are currently considered during the risk assessment process but may deserve more attention to provide additional protection.

3.4. Challenges in exposure assessment

Other limitations of epidemiological studies conducted on pesticides derive from uncertainty in exposure assessment. This represents a major limitation of studies on pesticides. Their specific limitations include the fact that most currently approved pesticides tend to have short elimination half-lives and that their use involves application of various formulations depending on the crop and season. As a result, accurate assessment needs to capture intermittent long-term exposure of these non-persistent chemicals as well as being able to quantify exposure to individual pesticides.

Numerous studies have assessed internal exposure by measuring urinary non-active metabolites common for a large group of pesticides (for example dialkyl phosphates for organophosphates, 3-phenoxybenzoic acid for pyrethroids or 6-chloronicotinic acid for neonicotinoids). These data may create uncertainty and should not be utilized to infer any risk because: a) a fraction of these metabolites might reflect direct exposure through ingestion of preformed metabolites from food and other sources, rather than ingestion of the parent compound; and b) the potency of the different parent pesticides can vary by orders of magnitude. Thereby, HBM data based on those urine metabolites can be unhelpful unless they are paired with other data indicating the actual pesticide exposure.

Ideally exposure should be quantified on an individual level using biomarkers of internal dose. As most available biomarkers reflect short term (few hours or days) exposure and given the cost and difficulty of collecting multiple samples over time, many studies quantify exposure in terms of external dose. Quantitative estimation of external dose needs to account for both frequency and duration of exposure and should preferably be done on an individual but no group level. Often external exposure is quantified using proxy measures such as:

- subject- or relative-reported jobs, job titles, tasks or other lifestyle habits which are being associated with the potential exposure to or actual use of pesticides in general and/or
- handling of a specific product or set of products and potential exposure to these as documented through existing pesticide records or diaries or estimated from crops grown;
- environmental data: environmental pesticide monitoring e.g. in water, distance from and/or duration of residence in a particular geographical area considered to be a site of exposure;

In many cases these proxy measures are recorded with use of questionnaires, which can be either interviewer-administered or based on self-report. The limitation here is that questionnaire data often rely on individual recall and knowledge and are thus potentially subject to both recall bias and bias introduced by the interviewer or study subjects. These sources of uncertainty can to some extent be quantified if the questionnaires are validated against biomarkers (that is, to what extent do individual questions predict biomarker concentrations in a sub-sample of participants). If the exposure is assessed retrospectively the accuracy of the recall is for obvious reasons more likely to be compromised and impossible to validate. When exposure is based on records, similar difficulties may occur due to e.g. incomplete or inaccurate records.

In many previous studies, duration of exposure is often used as a surrogate of cumulative exposure, assuming that exposure is uniform and continuous over time (e.g. the employment period) but this assumption must be challenged for pesticides. Although for some chemicals the exposure patterns may be fairly constant, exposures for many pesticides will vary with season, by personal protective equipment, and by work practices, and in many cases uses are not highly repetitive. At an individual level, exposures can vary on a daily and even hourly basis, and often involve several pesticides. This temporal variability can result in particularly high variation in systemic exposures for pesticides with short biological half-lives and considerable uncertainty in extrapolating single or few measurements to individual exposures over a longer term. Hence, many repeated measurements over time may be required to improve exposure estimates.

3.5. Inappropriate or non-validated surrogates of health outcomes

Reliance on clinically manifested outcomes can increase the likelihood that individuals who have progressed along the toxicodynamic continuum from exposure to disease but have not yet reached an overt clinical disease state will be misclassified as not having the disease (Nachman et al., 2011).

Thereby, delay in onset of clinical symptoms following exposure may cause underreporting where clinical assessment alone is used at an inappropriate point in time.

Surrogate outcomes may seem an attractive alternative to clinically relevant outcomes since there may be various surrogates for the same disease and they may occur sooner and/or be easier to assess, thereby shortening the time to diagnosis. A valid surrogate endpoint must however be predictive of the causal relationship and accurately predict the outcome of interest. Although surrogate markers may correlate with an outcome, they may not capture the effect of a factor on the outcome. This may be because the surrogate may not be causally or strongly related to the clinical outcome, but only a concomitant factor, and thus may not be predictive of the clinical outcome. The validity of surrogate outcomes may thus represent a major limitation to their use (la Cour et al., 2010).

Surrogate endpoints should thus be avoided unless they have been validated. Some criteria to assess the validity of a surrogate outcome include:

the surrogate has been shown to be in the causal pathway of the disease. This can be supported by the following evidence: correlation of biomarker response to pathology and improved performance relative to other biomarkers; biological understanding and relevance to toxicity (mechanism of response); consistent response across mechanistically different compounds and similar response across sex, strain and species; presence of dose-response and temporal relationship to the magnitude of response; specificity of response to toxicity; that is, the biomarker should not reflect the response to toxicities in other tissues, or to physiological effects without toxicity in the target organ.

at least one well conducted trial using both the surrogate and true outcome (Grimes and Schulz, 2005; la Cour et al., 2010). Several statistical methods are used to assess these criteria and if they are fulfilled the validity of the surrogate is increased. However, many times some uncertainty remains, making it difficult to apply surrogates in epidemiological studies (la Cour et al., 2010).

3.6. Statistical analyses and interpretation of results

The statistical analyses and the interpretation of scientific findings that appear in the epidemiologic literature on the relationship between pesticides and health outcomes do not substantially deviate from those reported in other fields of epidemiologic research. Therefore, the advantages and limitations of epidemiologic studies presented in section 2.5 also apply to the epidemiologic studies on pesticides.

The few distinctive features of the epidemiologic studies on pesticides include the following: a) sparse use of appropriate statistical analyses in the presence of measurement errors when assessing exposure to pesticides and b) paucity of information on other important factors that may affect the exposure-health relationship. These features are expanded on in the following paragraphs.

a) Statistical analyses in the presence of measurement errors

The difficulties inherent in correctly measuring exposure are frequent in many areas of epidemiologic research, such as nutritional epidemiology and environmental epidemiology. It is not easy to gauge the short- and long-term exposure outside controlled laboratory experimental settings. In large populations, individuals are exposed to a variety of different agents in a variety of different forms for varying durations and with varying intensities.

Unlike nutritional or environmental epidemiology, however, pesticide epidemiology has so far made little use of statistical analyses that would appropriately incorporate measurement errors, despite their wide availability and sizable literature on the topic. A direct consequence of this is that the inferential conclusions may not have been as accurate and as precise as they could have been if these statistical methods were utilized (Bengston et al., 2016; Dionisio et al., 2016; Spiegelman, 2016).

b) Information on other important factors of interest

Identifying and measuring the other relevant factors that might affect an outcome of interest is a recurrent and crucial issue in all fields of science. For example, knowing that a drug effectively cures a disease on average may not suffice if such drug is indeed harmful to children or pregnant women. Whether or not age, pregnancy, and other characteristics affect the efficacy of a drug is an essential piece of information to doctors, patients, drug manufacturers, and drug-approval agencies alike.

Pesticide epidemiology provides an opportunity for careful identification, accurate measuring and thorough assessment of possible relevant factors and their role in the exposure-health relationship. Most often, relevant factors have been screened as potential confounders. When confounding effects were detected, these needed to be adjusted for in the statistical analyses. This has left room for further investigations that would shed light on this important issue by reconsidering data that have already been collected and that may be collected in future studies. The statistical methods in the pesticide literature have been mainly restricted to standard applications of basic regression analyses, such as binary probability and hazard regression models. Potentially useful analytical approaches, such as propensity score matching, mediation analyses, and causal inference, does not seem to have been applied in pesticide epidemiology yet (Imbens and Rubin, 2015).

4. Proposals for refinement to future epidemiological studies for pesticide risk assessment

This chapter is aimed at addressing methods for assessment of available studies and proposals for improvement of such studies.

Most of the existing epidemiological studies on pesticides exposure and health effects suffer from a range of methodological limitations or deficiencies. Epidemiological studies would ideally generate semi-quantitative data or be able to have greater relevance to quantitative risk assessment with respect to the output from prediction models. This would allow epidemiological results to be expressed in terms more comparable to the quantitative risk assessments, which are more typically used in evaluating the risks of pesticides. The question arises how such epidemiological data could be considered for risk assessment when judged in comparison to the predictive models. A precisely measured quantitative dose-response relationship is presently extremely rarely attainable as a result of epidemiological studies.

The quality, reliability and relevance of the epidemiological evidence in relation to pesticide exposure and health effects can be enhanced by improving (a) the quality of each individual study and (b) the assessment of the combined evidence accrued from all available studies.

4.1. Assessing and reporting the quality of epidemiological studies

The quality and relevance of epidemiologic research should be considered when selecting epidemiological studies from the literature for use in risk assessment. The quality of this research can be enhanced by (Hernández et al., 2016; US-EPA, 2012):

a) an adequate assessment of exposure, preferentially biomarker concentrations at individual level reported in a way which will allow for a dose-response assessment;

b) a reasonably valid and reliable outcome assessment (well defined clinical entities or validated surrogates);

c) an adequate accounting for potentially confounding variables (including exposure to multiple chemicals); and

d) the conduct and reporting of subgroup analysis (e.g., stratification by gender, age, ethnicity).

It is widely accepted that biomedical research is subject to and suffers from diverse biases. Chalmers and Glasziou (2009) have estimated that approximately 85% of research investment in this area is